Faune-PACA Publication N°60

Le problème du collectionneur de coches





www.faune-paca.org
Le site des naturalistes de la région PACA



juin 2016

Le problème du collectionneur de coches

Mots clés : échantillonnage, problème du collectionneur de vignettes, coches

Auteur : Christian FRELIN

Citation : LPO PACA (2016). Le problème du collectionneur de coches. *Faune-PACA Publication n°60*, 9 pp.

christian.frelin@free.fr

Sommaire

Résumé

Plus on prospecte un peuplement et plus il est difficile d'observer de nouvelles espèces, comme si les espèces se raréfiaient. Pour cette raison, la richesse spécifique d'un peuplement ne peut jamais être définie. Cet article montre que cette raréfaction n'a pas forcément de base biologique. Elle peut être la conséquence d'un artefact d'échantillonnage bien connu de la théorie des probabilités et dénommé le « problème du collectionneur de coupons ».

Summary

The rate at which new species are detected in an area is well known to decrease as the sampling effort is increased. Cumulative numbers of species seem to reach an asymptote but never reach it. For this reason, species richnesses cannot be defined with precision. This paper shows that species rarefaction during sampling has no biological meaning. It arises as a consequence of a sampling artefact that relates to the well known coupon collector's problem of probability theory.

Remerciements

Je remercie Yvonne Delépine pour ses observations minutieuses et Cécile Lemarchand pour la mise en page.

1. Introduction

Peut-on définir le nombre d'espèces d'oiseaux que l'on peut rencontrer en France, dans le département des Alpes Maritimes, dans le Parc Naturel du Mercantour, dans la ville de Nice,... ? Ce type de question n'a pas de réponse (May, 1988, Gaston et Blackburn, 2000, Erwin, 1991, Mora et al., 2011). On a beau observer, compter et recompter, on bute invariablement sur la même difficulté. Plus on prospecte et plus il est difficile d'observer des nouvelles espèces. Mais la probabilité d'observer une nouvelle espèce, même faible n'est jamais nulle. L'observation d'un fou à Pieds rouges au lac de Sainte Croix dans les Alpes de Haute Provence en 2013 (Flitti et Rocha, 2014) en est une illustration spectaculaire. En conséquence, les inventaires ne sont jamais clos et la richesse spécifique, un paramètre essentiel qui caractérise chaque peuplement, n'est jamais connue avec précision. Les cocheurs font face au même problème. Plus leurs listes de coches s'allongent et plus il est difficile d'ajouter de nouvelles coches.

Le « problème du collectionneur de coupons » est un modèle classique de la théorie des probabilités (Ross 2012) qui s'applique aux situations où l'on échantillonne une population d'objets variés (coupons, figurines), jusqu'à obtenir le nombre total d'éléments présents. Cette note montre que le modèle s'applique à l'échantillonnage des populations d'oiseaux. Il permet de calculer l'effort de prospection qu'il faut faire pour observer un nombre donné d'espèces. Il explique aussi pourquoi chaque observateur croit percevoir une raréfaction progressive des espèces quand il cherche à allonger sa liste de coches.

2. Le problème du collectionneur de vignettes

Tous les enfants (et leurs parents) sont victimes un jour d'une stratégie de marketing, bien rodée et qui exploite leur envie de collectionner : objets cadeau, vignettes « Panini » à l'effigie des joueurs de la coupe du monde de Football, cartes « Pokemon », figurines de toutes sortes, etc. Les vignettes « Panini » par exemple, sont achetées par paquets de cinq vignettes et collées dans un album « Collector ». Le jeu consiste, moyennant finances, à obtenir toutes les effigies de la collection et à compléter l'album. Au début, tout se passe bien. Chaque lot de vignettes apporte de nouvelles effigies. Puis la source semble se tarir ; on achète de nouvelles vignettes, mais on tombe toujours sur des mêmes effigies. On se demande alors si le jeu n'est pas « pipé » et si certaines effigies n'ont pas été raréfiées volontairement pour promouvoir les ventes. Puis, surprise, on se rend compte que le fils du voisin fait la même collection et qu'il possède déjà les vignettes « rares » que l'on cherche. Tout se passe comme si les vignettes qui étaient rares pour l'un ne l'étaient pas pour l'autre. On échange alors les doublons, les albums se remplissent rapidement et on peut se tourner vers un nouveau jeu. La question que tout parent devrait se poser avant de s'engager dans une telle collection est :

Combien faudra-t-il acheter de vignettes pour compléter l'album ? Et : quel sera le coût final ?

Ainsi posé, il s'agit d'un problème classique de théorie des probabilités connu sous le nom de « Problème du collectionneur de vignettes ». Nous considérerons ici le cas le plus simple, celui ou les vignettes sont vendues séparément et sont équiréparties. Les cas, plus complexes, ou les vignettes sont vendues par lots (par exemple 5 dans le cas des vignettes « Panini ») et ou les vignettes ne sont pas équiréparties sont traités par ailleurs (Sardy et Velenik, 2011, Ferrante et Saltalamacchia, 2014).

La démonstration est assez simple mais elle est peu intuitive. On considère une collection de n vignettes différentes qui représentent les effigies des 660 joueurs engagés dans la Coupe du Monde. Les vignettes sont produites en grandes quantités. Toutes les effigies sont également représentées. Le collectionneur achète les vignettes une à une, sans choisir l'effigie qu'il achète.

La première vignette achetée est nouvelle, bien sûr ; le collectionneur la colle immédiatement dans son album. A la seconde vignette, deux solutions sont possibles

1. Elle est identique à la première. La probabilité pour qu'elle le soit est égale à 1/n.

2. Elle est différente de la première. La probabilité pour qu'elle le soit est la probabilité complémentaire (1-1/n), soit (n-1)/n.

Si k effigies différentes ont déjà été obtenues et collées dans l'album, la probabilité d'obtenir une nouvelle effigie au tirage (achat) suivant est : (n-k)/n

Lorsqu'on a obtenu (n-1) effigies différentes, c'est à dire lorsqu'il ne reste plus qu'une seule effigie à obtenir, la probabilité de l'obtenir au tirage suivant est 1/n. Autrement dit, la probabilité de tirer une effigie quelconque est 1/n à tout moment. Mais la probabilité d'avoir une « nouvelle » effigie dépends du nombre de tirages. Elle est égale à 1 au premier tirage. Elle diminue ensuite selon une progression géométrique de raison (n-k+1)/n et atteint 1/n quand il ne reste qu'une seule effigie à acquérir

Le nombre total de vignettes qu'il faudra acheter pour compléter un album de n effigies est donné, en moyenne,

par la relation

n * (1 +1/2 + 1/3 +… + 1/n), soit

n * (ln(n) + γ) vignettes (Relation 1).

ou ln est le logarithme népérien et γ est la constante d'Euler-Mascheroni (γ = 0,577)

La collection « Panani » éditée pour les mondiaux de 2010 comportait 660 effigies. La relation (1) permet de calculer qu'il fallait acheter, en moyenne, 4666 vignettes pour obtenir toutes les effigies, soit 7 fois plus de vignettes que n'en comportait la collection complète des effigies. Soit aussi un coût de 933 euros si chaque vignette était vendue 20 centimes.

Le cas où les vignettes sont vendues par paquets de 5 a été analysé par Sardy et Velenik, (2011). Les résultats sont très similaires. Il faut acheter en moyenne 931 pochettes à 1 euro (soit 4655 vignettes et 931 euros) pour compléter la collection. Sardy et Velenik (2011) ont analysé ensuite comment l'acquisition des nouvelles effigies ralentit avec le temps. Les premières 1166 vignettes fournissent en moyenne 547 effigies distinctes. Les 1166 vignettes suivantes apportent 94 effigies nouvelles. Les 1166 vignettes suivantes n'en apportent qu'environ 16. Enfin il faut acheter encore1168 vignettes pour obtenir les 3 dernières effigies.

3. Le problème du collectionneur de coches

Collectionner des vignettes ou des coches revient au même. Les oiseaux observés correspondent aux vignettes ; les espèces cochées correspondent aux effigies. Prenons par exemple un peuplement de 100 espèces également réparties (équiréparties). L'observateur est posté et il attend. Les oiseaux se présentent un à un et l'espèce est identifiée. La question est : Combien faut-il observer d'oiseaux pour arriver à cocher les 100 espèces du peuplement ? La réponse est donnée par la relation (1)

100 * (ln(100) + 0,577) soit 518 observations, c'est à dire 5,2 fois plus que le nombre d'espèces présentes

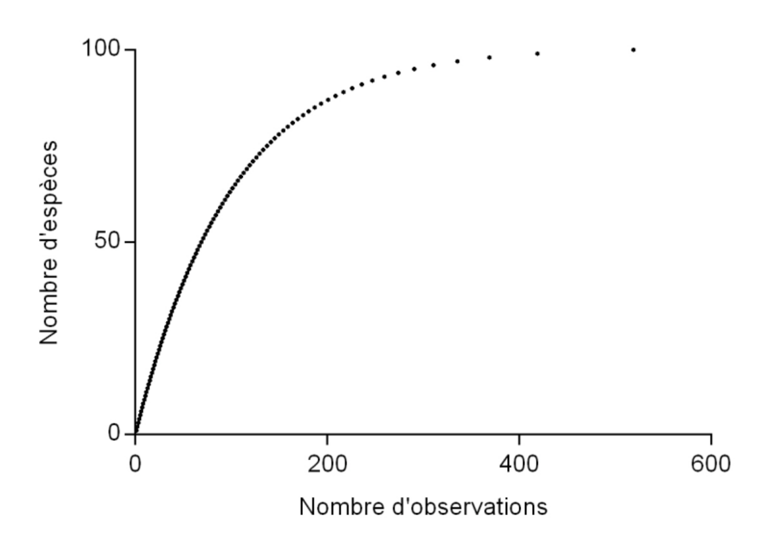


Figure 1 : Influence de l'effort de prospection sur le nombre d'espèces identifiées. On a supposé que le peuplement comportait 100 espèces équiréparties. Les dernières espèces recensées semblent plus rares. Cette impression est fausse car les espèces sont équiréparties.

La Figure 1 montre comment le nombre des coches augmente en fonction de l'effort de prospection. Les nouvelles espèces s'accumulent rapidement au début de la période d'observation, puis elles se raréfient. On retrouve un fait bien connu. Plus on prospecte et plus il est difficile d'observer une espèce nouvelle, comme si les espèces devenaient de plus en plus rares.
Combien d'espèces peut-on observer dans les Alpes maritimes, à Nice,….. ? La réponse est décevante, mais elle est claire : cela dépend du temps passé à prospecter. Plus ce temps est long et plus le nombre d'espèce sera grand.

4. La notion de rareté

L'interprétation habituelle des courbes de saturation (Figure 1) est la suivante. Le peuplement comporte des espèces communes qui sont rapidement détectées et sont identifiées dès le début de l'observation. Les autres espèces sont plus rares et sont détectées plus tard. La dernière espèce cochée, celle qui a demandé le plus d'effort de prospection (ou l'attente la plus longue) est la plus rare. Cette impression de rareté n'est pas réelle car nous avions supposé au départ que toutes les espèces étaient également représentées. La difficulté provient de la confusion entre « probabilité d'observer une espèce quelconque » et « probabilité d'observer une nouvelle espèce ». Avec n espèces (ici n = 100), la probabilité d'observer une espèce quelconque est égale à 1/n. Cette valeur ne change pas pendant la prospection et elle est la même pour toutes les espèces (le peuplement est

équiréparti). Par contre, la probabilité d'observer une espèce **nouvelle** change avec le temps. Elle est égale à 1 quand la première espèce se présente à l'observateur. Puis elle diminue et elle atteint 1/n à la dernière coche. Ceci donne l'impression, fausse, que les dernières espèces cochées étaient plus rares que les premières. Selon Sardy et Velenik, (2011), l'impression de rareté qui subsiste alors même que les vignettes sont distribuées de façon équiprobable est de nature « psychologique ». Elle traduit la mauvaise intuition du hasard par nos cerveaux.

Inventorier un peuplement revient à identifier et cocher toutes les espèces présentes. Inventorier complètement un peuplement de 10 espèces équiréparties demande, en moyenne, 28 observations. La probabilité de noter une espèce nouvelle passe de 1 à 0,1 (1/10) après 27 observations. Inventorier un peuplement de 100 espèces équiréparties demande, en moyenne 518 observations. La probabilité de cocher une espèce nouvelle passe de 1 à 0,01 (1/100) après 517 observations

Ainsi, si n = 10, la probabilité de détecter la dernière espèce du peuplement est 10 fois plus grande que la probabilité de détecter la dernière espèce d'un peuplement de 100 espèces. La conséquence est simple. Les peuplements les plus riches sembleront comprendre plus d'espèces rares que les peuplements paucispécifiques.

De la même manière, les observateurs les plus chevronnés (les « anciens » qui ont accumulé le plus de coches) pourront avoir l'impression (fausse) que les avifaunes s'appauvrissent.

5. Application

Le « Chemin de la Forna » est un lieu-dit situé en zone résidentielle sur la commune de La Turbie (06) au nord de Monaco. Le lieu-dit est situé sur un axe migratoire des Alpes Maritimes : le fort de la Revère (Clot et Lemarchand, 2012). 3112 observations ont été réalisées par Yvonne Delépine entre le 14/02/2010 et le 22/11/2014. Les observations sont étalées sur 433 jours de prospection répartis régulièrement au cours du cycle annuel. 73 espèces ont été observées.

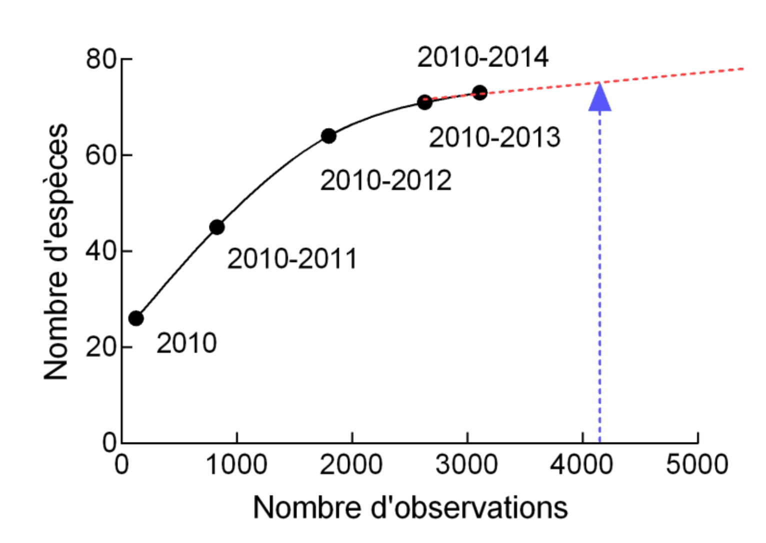


Figure 2 : Évolution du nombre cumulé d'espèces observées au lieu-dit « Chemin de La Forna » entre 2010 et 2014. La ligne pointillée indique la courbe obtenue si l'on extrapolait les deux derniers points. Les données ont été extraites de Faune Paca (http://www.faune-paca.org). Une « observation » est définie comme l'observation d'un ou plusieurs individus d'une espèce par jour de prospection. C'est le format d'entrée habituel dans Faune Paca. Il s'agit donc de relevés en présence/absence. Noter que les abscisses sont de observations et pas le nombre d'années de prospection

La prospection a-t-elle été suffisante pour identifier toutes les espèces qui fréquentent le lieu-dit ? L'approche classique consiste à tracer une courbe de saturation (Ferry, 1976, Blondel 1979). La Figure 2 montre que la richesse cumulée a augmenté rapidement au cours des trois premières années, puis elle semble se stabiliser. En extrapolant la courbe à partir des deux derniers points, on montre que 1000 observations supplémentaires apporteraient 4 espèces nouvelles. À titre de comparaison, les 1000 premières observations avaient détecté 45 espèces !

Le point important ici est qu'il a fallu identifier 3112 oiseaux pour obtenir la liste des 73 espèces qui fréquentent le lieu-dit. Le modèle du collectionneur de vignettes prédisait 9 fois moins (355 observations). La différence indique que l'échantillonnage du peuplement du chemin de la Forna était biaisé par la présence d'espèces plus rares que les autres.

Le modèle du collectionneur de coupons s'applique aux situations ou toutes les espèces sont également représentées, c'est à dire aux peuplements les plus diversifiés. Il fournit une hypothèse nulle pour analyser la structure d'un peuplement.

6. Conclusion

Les collectionneurs de coches existent depuis très longtemps, bien plus longtemps que les collectionneurs de vignettes « Panini ». Pour des raisons d'antériorité, le « problème du collectionneur de vignettes » pourrait être renommé « problème du collectionneur de coches ».

Bibliographie

Blondel, J. (1979) Biogéographie et écologie. Masson ed, 173 p

Ferrante, M. et Saltalamacchia, M. (2014). The coupon collector's problem. http://mat.uab.cat/matmat/PDFv2014/v2014n02.pdf

Ferry, C. (1976) Un test facile pour savoir si la richesse mesurée d'un peuplement se rapproche de sa richesse réelle. Le Jean le Blanc 15, 21-28.

Flitti, A. et Rocha, E. (2014) Première mention française du Fou à pieds rouges. Ornithos, 21, 158-161.

Erwin T. L (1991) How many species are there? Revisited. Conserv Biol 5: 1–4.

Gaston K, Blackburn T (2000) Pattern and process in macroecology. Blackwell Science Ltd.

Clot, T. & Lemarchand, C. (2012), La migration postnuptiale des oiseaux au fort de la Revère en 2012 (Alpes Maritimes). Faune-PACA Publication, 19 : 128 pp

May R. M (1988) How many species are there on earth? Science, 241: 1441–1449.

Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B (2011) How Many Species Are There on Earth and in the Ocean? PLoS Biol 9(8): e1001127.

Ross, S. 2012. A first course in probability. 9th Edition, Pearson.

Sardy, S. et Velenik, Y. (2010) Petite collection d'informations utiles pour collectionneur compulsif. http://images.math.cnrs.fr/Petite-collection-d-informations.html

La faune de la région PACA

Le territoire de la région Provence-Alpes-Côte d'Azur est le plus riche et le plus diversifié en nombre d'espèces en France métropolitaine. La région PACA abrite 245 espèces d'oiseaux nicheurs sur 275 espèces recensées en France, 70 % des 143 espèces de mammifères, 80 % des 34 Reptiles, 61 % des 31 Amphibiens, 85 % des 240 papillons de jour et 74 % des 100 libellules.

Le projet www.faune-paca.org

En mars 2016, le site http://www.faune-paca.org a dépassé le seuil des **4 millions de données** portant sur les oiseaux, les mammifères, les reptiles, les amphibiens, les libellules et les papillons diurnes. Ces données zoologiques ont été saisies et cartographiées en temps réel. Le site http://www.faune-paca.org s'inscrit dans une démarche collaborative et mutualiste de mise à disposition d'un atlas en ligne actualisé en permanence. Faune-paca.org est un projet développé par la LPO PACA et consolidé au niveau national par le réseau LPO sur le site www.ornitho.fr.

Ce projet est original et se caractérise par son rôle fédérateur, son efficacité, sa fiabilité, son ouverture aux professionnels de l'environnement et aux bénévoles. Chacun est libre de renseigner les données qu'il souhaite, de les rendre publiques ou non, et d'en disposer pour son propre usage comme bon lui semble. Il est modulable en fonction des besoins des partenaires. Il est perpétuellement mis à jour et les données agrégées sont disponibles sous forme de cartographies et de listes à l'échelle communales pour les acteurs du territoire de la région PACA.

Faune-PACA Publication

Cette nouvelle publication en ligne Faune-PACA publication a pour ambition d'ouvrir un espace de publication pour des synthèses à partir des données zoologiques compilées sur le site internet éponyme www.faune-paca.org. Les données recueillies sont ainsi synthétisables régulièrement sous forme d'ouvrages écrits de référence (atlas, livres rouges, fiches espèces, fiches milieux, etc.), mais aussi, plus régulièrement encore, sous la forme de publications distribuées électroniquement. Faune-PACA Publication est destiné à publier des comptes-rendus naturalistes, des rapports d'études, des rapports de stage pour rythmer les activités naturalistes de la région PACA. Vous pouvez soumettre vos projets de publication à Olivier Hameau, rédacteur en chef de la publication olivier.hameau@lpo.fr et à Amine Flitti, responsable des inventaires et administrateur des données sur faune-paca.org amine.flitti@lpo.fr.

Faune-PACA Publication n°60

LPO — AGIR pour la BIODIVERSITÉ Provence-Alpes-Côte d'Azur

Article édité par la
LPO PACA
Villa Saint-Jules
6, avenue Jean Jaurès
83400 HYERES
tél: 04 94 12 79 52 Fax: 04 94 35 43 28
Courriel: paca@lpo.fr
Web: http://paca.lpo.fr

Directeur de la publication : Benjamin KABOUCHE
Rédacteur en chef : Olivier HAMEAU
Comité de lecture du n°60 : Amine FLITTI, Cécile LEMARCHAND
Administrateur des données www.faune-paca.org : Amine FLITTI.
Photographies couverture : Observateurs à la Revère © Eve Lebegue, Fou à pied rouge © Frank Dhermain, Vignette Zinedine ZIdane © oldschool Panini

ISSN en cours

